Matematiske emner

Usandsynlige sandsynligheder

Ole Witt-Hansen 2009

Indhold

1. Fødselsdagsproblemet

Der findes som bekendt i sandsynlighedsregningen flere eksempler på, at en udregning giver et overraskende resultat. Et af mest kendte er nok fødselsdagsproblemet, som fandtes i Kristensen og Rindungs lærebog i sandsynlighedsregning indtil 1988 – inklusive en udledning! Udledningen er gennemført nedenfor i min egen formulering.
Problemet, som findes beskrevet i flere lærebøger i sandsynlighedsregning er følgende:

Hvis man har en klasse på 24 elever, hvad er så sandsynligheden for at mindst to elever har fødselsdag på samme dag. Mere generelt, hvis man har n personer, hvor $n < 365$, hvad er så sandsynligheden for, at mindst to har fødselsdag på samme dag.

For tilfældet $n = 24$ er det interessant (f.eks. med henblik på et væddemål), hvorvidt denne sandsynlighed er større end ½.

Vi bestemmer først antallet af elementer i udfaldsrummet. Da hver person har 365 muligheder for fødselsdag er der

$$365\cdot365\cdot365\cdot\ldots\cdot365 = 365^n$$

muligheder for placering af fødselsdage på de n personer.

I stedet for at udregne den ønskede sandsynlighed, udregner vi i stedet sandsynligheden $P(H)$ for den komplementære hændelse, altså sandsynligheden for, at de n personer alle har fødselsdag på forskellige dage. Sandsynligheden for mindst et sammenfald, kan så udregnes som 1- $P(H)$.
Antallet af gunstige udfald for H er:

(1.1) $$365\cdot364\cdot363\cdot\ldots\cdot(365-n+1) \quad \text{(n - faktorer)}$$

Der er 365 muligheder for det første valg, 364 for det andet osv. $P(H)$ udregnes da som altid, som antallet af gunstige udfald divideret med antallet af mulige udfald.

(1.2) $$P(H) = \frac{365\cdot(365-1)\cdot(365-2)\cdot\ldots(365-n+1)}{365\cdot365\cdot365\cdot365\cdot\ldots\cdot365} = \frac{365}{365}\cdot\frac{365-1}{365}\cdot\frac{365-2}{365}\cdot\ldots\cdot\frac{365-n+1}{365}$$

Med lidt tålmodighed, kan man på en lommeregner udregne dette udtryk for $n = 24$. Man finder at sandsynligheden er lidt mindre end ½, og dermed, at der (lidt overraskende) er mere end 50% chance for sammenfald af fødselsdage i en klasse med 24 elever.
For n = 22 er der derimod mindre end 50% chance for sammenfald af fødselsdage.

Vi vil nu vise, hvorledes man simplere, og med god tilnærmelse kan beregne sandsynligheden $P(H)$.

For at udlede en formel dividerer vi 365 op i tælleren i hver af brøkerne i udtrykket ovenfor, og tager den naturlige logaritme til begge sider.

(1.3)
$$P(H)=\frac{365}{365}\cdot\frac{365-1}{365}\cdot\frac{365-2}{365}\cdot\ldots\frac{365-n+1}{365}=1\cdot(1-\frac{1}{365})\cdot(1-\frac{2}{365})\cdot\ldots\cdot(1-\frac{n-1}{365})$$
$$\ln P(H)=\ln(1-\frac{1}{365})+\ln(1-\frac{2}{365})+\ldots+\ln(1-\frac{n-1}{365})$$

For den naturlige logaritmefunktion, gælder der følgende tilnærmelsesformel:

$$ln(1+h) = h,$$

når h er numerisk meget mindre end 1 (mindre end 1/10).

Anvendes denne formel på alle de n-1 led fås:

$$\ln P(H)=-\frac{1}{365}-\frac{2}{365}-\frac{3}{365}-\ldots\frac{n-1}{365}=-\frac{1}{365}(1+2+3+\ldots(n-1))$$
$$\ln P(H)=-\frac{1}{365}\frac{n(n-1)}{2}$$

Vi søger da at bestemme n således at:

$$P(H) < ½ \Leftrightarrow ln\ P(H) < ln½ \Leftrightarrow ln\ P(H) < -0,6931$$

Dette giver uligheden

$$-\frac{1}{365}\frac{n(n-1)}{2}<-0{,}6931 \quad\Leftrightarrow\quad n(n-1)>505{,}963$$

Hvilket er en 2. grads ulighed:

$$n^2-n-505{,}963>0 \quad\Leftrightarrow\quad x<-22 \quad\vee\quad x>23$$

Idet den tilsvarende 2.gradsligning har løsningerne:

$$n^2-n-505{,}963=0 \quad\Leftrightarrow\quad n=\frac{1\pm 45}{2}=\left(\begin{array}{l}23\\-22\end{array}\right.$$

Altså, når antallet n er større end eller lig med 23, er der mindre end 50% chance for at alle har fødselsdag på forskellige dage – eller
For n >23 der er mere end 50% chance for, at mindst to har fødselsdag på samme dag.

2. Mønten i de tre æsker - problemet

En formulering af problemet er, at man har tre identiske æsker. I den ene æske er en mønt. Spillet går ud på at vælge æsken med mønten.

Man vælger én af æskerne (uden at åbne). Derpå fjernes én af de to øvrige æsker, som ikke indeholder mønten.
Man tilbyder spilleren at vælge om mellem de to resterende æsker.

Er det en fordel at vælge om? Svaret er altid ja.

Dette er ikke noget egentligt kombinatorisk problem, men blot formuleret på en måde, der kan forvirre. At elever har svært ved at gennemskue det, er forståelig nok, men at matematiklærere kan diskutere det i længere tid, og med indlæg i blade er mig komplet ubegribeligt.

Den første æske er valgt med sandsynligheden 1/3. Sandsynligheden for at mønten befinder sig i en af de to resterende æsker er derfor 2/3. Disse sandsynligheder kan der ikke laves om på, når valget er truffet. Da den ene af de to resterende æsker, som *ikke* indeholder mønten fjernes, må sandsynligheden for at mønten befinder sig i den resterende æske være 2/3.
Derfor skal man altid vælge om. Sandsynligheden for at vælge æsken med mønten, vil være dobbelt så stor ved omvalg.

3. Antallet af permutationer hvor intet element svarer til sig selv

I forbindelses med børnekortspillet "krig", blev jeg engang (som barn) gjort opmærksom på et lignende problem som fødselsdagsproblemet. Hvis to spillere har hver sit spil kort, og hvis de samtidig tager et kort fra bunken og sammenligner, så kan man indgå et væddemål om hvorvidt, der vil være et sammenfald af kort i de 52 (eller 13) forsøg.
Ens fornemmelse vil nok gå i retning af, at sandsynligheden for sammenfald er mindre end ½.

Formuleret på en anden måde: Er antallet af permutationen af n elementer, hvor intet element svarer til sig selv større eller mindre en halvdelen af mulige permutationer $n!$

Dette kræver så en bestemmelse af antallet af permutationer $Q(n)$, hvor intet element svarer til sig selv. Dette viser sig imidlertid at være langt vanskeligere at afgøre end fødselsdagsproblemet. Kombinatoriske problemer kan være svære at overskue, og bestemmelsen af $Q(n)$ er i den henseende et "glimrende" eksempel.

Som vist nedenfor konvergerer sandsynligheden for at ingen elementer svarer til sig selv mod $1/e = 0{,}3679$ for n gående mod uendelig, hvilket for kortspillet, hvor man på tilfældig måde trækker et kort fra to identiske bunker, giver sandsynligheden 0,6321 for sammenfald – altså en relativ stor overvægt.

3.1 Opstilling af en rekursionsformel for $Q(n)$

Det er imidlertid ret ligetil at lave en rekursionsformel for $Q(n)$, hvor $Q(n)$ er udtrykt ved alle de forrige $Q(n\text{-}1)$, $Q(n\text{--}2)$, ..., $Q(2)$

Antallet af permutationer af n (forskellige) elementer betegnes som sædvanlig $P(n,n) = n!$
Antallet af permutationer, hvor der en k ens elementer er:

$$P_k(n,n) = \frac{n!}{k!} \ ,$$

idet det er antallet af permutationer, divideret med antallet af permutationer af de k ens elementer.

Antallet af q kombinationer af en n mængde, skrives som sædvanlig:

(3.1) $$K(n,q)=\binom{n}{q}=\frac{n!}{(n-q)!q!}$$

$Q(n)$ betegner så antallet af permutationer af n elementer, hvor intet element svarer til sig selv.

Vi opstiller nu en rekursionsformel på følgende måde. Fra samtlige permutationer $n!$ subtraherer vi de permutationer, hvor netop ét element svarer til sig selv, de permutationer, hvor netop to elementer svarer til sig selv,… ,de permutationer, hvor alle elementer svarer til sig selv.
På denne måde får vi rekursionsformlen.

(3.2) $$Q(n)=n!-\binom{n}{1}Q(n-1)-\binom{n}{2}Q(n-2)-....-\binom{n}{n-2}Q(2)-1$$

I det tredje led f.eks. kan de to elementer som svarer til sig selv vælges på $K(n,2)$ forskellige måder. De øvrige n-2 elementer kan permuteres på $Q(n$-2) måder, så intet element svarer til sig selv.

Formlen kan ikke anvendes for n = 2, men her ved vi, at svaret er $Q(2) = 1$, nemlig når elementerne byttes om. Vi finder herefter:

$Q(3) = 3! – 3\cdot Q(2) - 1 = 2$

$Q(4) = 4! - 4\cdot Q(3) - 6\cdot Q(2) - 1 = 9$

$Q(5) = 5! - 5\cdot Q(4) -10\cdot Q(3) - 10\cdot Q(2)-1=5! - 5\cdot 9-10\cdot 2- 10\cdot 1-1= 120 – 45 – 20 – 10 -1 = 44$.

Det ville selvfølgelig være mere tilfredsstillende, at have en formel for $Q(n)$, som ikke var en rekursionsformel. Det er som tidligere nævnt ikke så ligetil ad analytisk vej.

Hvis vi i rekursionsformlen ovenfor isolerer $n!$, kan formlen skrives:

(3.3) $$n!=\sum_{k=0}^{n}\binom{n}{k}Q(n-k)$$

Hvis vi dividerer med $n!$ får man derefter:

(3.4) $$1=\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}Q(n-k)=\sum_{k=0}^{n}\frac{Q(n-k)}{k!(n-k)!}=\sum_{k=0}^{n}\frac{Q(k)}{(n-k)!k!}$$

I det følgende får vi brug for denne formel, formlen for en uendelig kvotientrække med kvotienten x, samt rækkeudviklingen for e^x.

(3.5) $$\sum_{k=0}^{\infty}x^k=\frac{1}{1-x} \quad \text{for } -1<x<1 \qquad e^x=\sum_{k=0}^{\infty}\frac{x^k}{k!}$$

Vi ser da på udtrykket:

(3.6) $$e^x\left(\sum_{q=0}^{\infty}\frac{Q(q)}{q!}x^q\right)=\left(\sum_{k=0}^{\infty}\frac{x^k}{k!}\right)\left(\sum_{q=0}^{\infty}\frac{Q(q)}{q!}x^q\right)=R(x)\left(\sum_{k=0}^{\infty}\frac{x^k}{k!}\right)$$

Den første faktor er e^x. Den anden faktor betegnes $R(x)$. Når vi ganger de to (uendelige summer) sammen, samler vi alle led hvor $k + q = n$, så alle potenserne $x^{k+q} = x^n$.
Samtlige led kan da skrives, hvor k=0,1..,n

(3.7) $$\sum_{k=0}^{n}\frac{Q(n-k)}{k!(n-k)!}x^n = x^n\sum_{k=0}^{n}\frac{Q(n-k)}{k!(n-k)!}=x^n$$

på grund af betingelsen: $\sum_{k=0}^{n}\frac{Q(k)}{(n-k)!k!}=1$

Gentager vi dette for alle leddene n=0, n=1,……kan produktet af de to summer skrives:

$$\sum_{k=0}^{\infty}x^k=\frac{1}{1-x}.$$

Vi har altså:

(3.8) $$e^xR(x)=e^x\sum_{q=0}^{\infty}\frac{Q(q)}{q!}x^q=\frac{1}{1-x}\qquad eller\qquad \sum_{n=0}^{\infty}\frac{Q(n)}{n!}x^n=\frac{e^{-x}}{1-x}$$

Vi gange over med 1- x og sammenligner derefter led på begge sider med samme potens af x

$$\sum_{n=0}^{\infty}\frac{Q(n)}{n!}x^n=\frac{e^{-x}}{1-x}\qquad\Leftrightarrow$$

$$(1-x)\sum_{n=0}^{\infty}\frac{Q(n)}{n!}x^n=e^{-x}\qquad\Leftrightarrow\qquad(1-x)\sum_{n=0}^{\infty}\frac{Q(n)}{n!}x^n=\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}x^n$$

(3.9) $$\sum_{n=0}^{\infty}\frac{Q(n)}{n!}x^n-\sum_{n=0}^{\infty}\frac{Q(n)}{n!}x^{n+1}=\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}x^n$$

Sammenlignes leddene med samme eksponent n, får man:

$$\frac{Q(n)}{n!}-\frac{Q(n-1)}{(n-1)!}=\frac{(-1)^n}{n!}\qquad\Leftrightarrow\qquad\frac{Q(n)}{n!}=\frac{Q(n-1)}{(n-1)!}+\frac{(-1)^n}{n!}\qquad\Leftrightarrow$$
$$Q(n)=nQ(n-1)+(-1)^n$$

Idet $Q(1)=0$, finder man $Q(2)=1$, $Q(3)=2$; $Q(4)=9$, $Q(5)=44$, $Q(6)=265$;

Endvidere ses at sandsynligheden $P(n)$ for at intet element svarer til sig selv er:

$$\frac{Q(2)}{2!}=\frac{Q(1)}{1!}+\frac{(-1)^2}{2!}=\frac{(-1)^2}{2!}=\frac{1}{2}$$

$$\frac{Q(3)}{3!} = \frac{Q(2)}{2!} + \frac{(-1)^3}{3!} = \frac{(-1)^2}{2!} + \frac{(-1)^3}{3!}$$
$$\frac{Q(4)}{4!} = \frac{Q(3)}{3!} + \frac{(-1)^4}{4!} = \frac{(-1)^2}{2!} + \frac{(-1)^3}{3!} + \frac{(-1)^4}{4!}$$

$$P(n) = \frac{Q(n)}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

hvoraf ses, at

$$P(n) \to \frac{1}{e} \quad for \quad n \to \infty$$

Sandsynligheden for at intet element svarer til sig selv ved en permutation af n elementer (for $n>4$) med tilnærmelse $1/e = 0{,}368$.
Hvis vi beregner sandsynligheden for $n = 1..5$

$$P(2) = \frac{Q(2)}{2!} = \tfrac{2}{6} = 0{,}50 \ , \quad P(3) = \frac{Q(3)}{3!} = \tfrac{2}{6} = 0{,}33$$
$$P(4) = \frac{Q(4)}{4!} = \tfrac{3}{8} = 0{,}375 \ , \quad P(5) = \frac{Q(5)}{5!} = \tfrac{44}{120} = 0{,}367$$

Som man ser så er resultatet allerede for $n = 4$, meget tæt på 1/e = 0,368.

Rekursionsformlen $Q(n) = nQ(n-1) + (-1)^n$ for antallet af permutationer, hvor intet element svarer til sig selv er først udledt direkte af de Montmort i 1708. Senere er vist, at denne formel fører til rækkeformlen ovenfor. Argumentationen er imidlertid lidt tricky.

Vi deler permutationerne af $1..n$ op efter det k'te element, hvor $k = 2,3..n$.
Først ser vi på de permutationer uden fixpunkter, hvor elementet "1" anbringes på den k'te plads. De resterende n-1 elementer kan permuteres på $Q(n-1)$ forskellige måder uden fixpunkter.

Dernæst ser vi på de permutationer, hvor "1" ikke står på den k'te plads, men altså på pladserne $2..k-1, k+1..n$ og k står på den første plads. De resterende n-2 elementer, kan permuteres på $Q(n-2)$ forskellige måder uden fixpunkter. Idet k antager værdierne $2..n$, er der n-1 muligheder for k.

Man kan da indse, at vi får antallet af samtlige permutationer uden fixpunkter ved formlen.

$$Q(n) = (n-1)(Q(n-1)+Q(n-2))$$

Denne rekursionsformel kan imidlertid forsimples til

$$Q(n)=nQ(n-1)+(-1)^n$$

Dette bevises ved induktion.
Formlen er rigtig for n=2. Vi antager, at den er rigtig for n, og viser, at den også er rigtig for n+1.
Vi indsætter da

$$Q(n)=nQ(n\text{-}1)+(-1)^n \quad \Leftrightarrow \quad nQ(n\text{-}1)= Q(n) -(-1)^n$$

på højre side af den første rekursionsformel med $n+1$.

$$\begin{aligned} Q(n+1) &=(n+1)(Q(n)+Q(n\text{-}1))= \\ &=nQ(n)+ nQ(n\text{-}1)= nQ(n)+ Q(n) -(-1)^n = (n+1)Q(n)+(-1)^{n+1} \end{aligned}$$

Dividerer vi den nye forsimplede rekursionsformel med $n!$, finder vi sandsynligheden for at få en permutation uden fixpunkter.

$$P(n) = \frac{Q(n)}{n!} = \frac{Q(n-1)}{(n-1)!} + \frac{(-1)^n}{n!}$$

Man ser at der gælder

$$P(n) - P(n-1) = \frac{Q(n)}{n!} - \frac{Q(n-1)}{(n-1)!} = \frac{(-1)^n}{n!}$$

$$P(1) = 0,\ P(2) = \tfrac{1}{2!}, \quad P(3) = \tfrac{1}{2!} - \tfrac{1}{3!},\ \ P(4) = \tfrac{1}{2!} - \tfrac{1}{3!} + \tfrac{1}{4!}$$

$$P(n) = \tfrac{1}{2!} - \tfrac{1}{3!} + \tfrac{1}{4!} + \ldots + \tfrac{(-1)^n}{n!}$$

$$P(\infty) = e^{-1}$$

Svarende til den anførte formel ovenfor.

4. St. Petersborg paradokset

St. Petersborg paradokset er først fremsat af Daniel Bernoulli i forbindelse med hasard spil. Spillet er et eksempel, hvor den matematisk beregnede middelgevinst er uendelig stor, men hvor ingen rationelle mennesker alligevel ikke ville drømme om at investere en stor kapital i spillet. Sådanne paradokser findes også i økonomisk teori.

Selv om en uendelig kapital ikke findes i virkeligheden, kan man alligevel godt drage lignende konklusioner, hvis der er et loft M over gevinsten.

Spillet er i øvrigt meget simpelt: Man spiller med en mønt (eller sort/rød på en roulette). Spilleren indgår et væddemål med banken, hvor spilleren som indsats deponerer S enheder. S bestemmes af banken.

Hvis spilleren slår krone N gange i træk, vinder han 2^N enheder.

Sandsynligheden for en sekvens af møntkast med N krone efterfulgt af plat, er:

$$P(N) = (\tfrac{1}{2})^N \, \tfrac{1}{2} = (\tfrac{1}{2})^{N+1}$$

Og gevinsten er følgelig: $G(N) = 2^N - S$. Middelgevinsten bliver herefter:

$$E(G)=\sum_{N=0}^{\infty}P(N)G(N)=\sum_{N=0}^{\infty}(\tfrac{1}{2})^{N+1}(2^N-S)=\sum_{N=0}^{\infty}(\tfrac{1}{2}-(\tfrac{1}{2})^{N+1}S)=\infty-S\tfrac{1}{2}\frac{1}{1-\frac{1}{2}}=\ \infty-S=\infty$$

Det er lidt overraskende, at middelgevinsten altid vil være uendelig stor, uafhængig af indsatsen. Et Casino, der spiller mange spil, vil ikke risikere at få sprængt banken, med mindre at indsatsen S er meget høj, i hvilket tilfælde ingen vil spille spillet. Det er paradokset.

Paradokset ligger nu ikke i den matematiske udregning, men i den asymmetri, der er mellem et Casino på en ene side og f.eks. 10.000 spillere på den anden. Et Casino bærer alene gevinsten eller tabet, mens der blandt spillerne måske kun er én, som opnår en stor gevinst, mens de øvrige 9.999 vil tabe.
Hvis de 10.000 spillere slog sig sammen om at dele den samlede gevinst, ville spillet være fordelagtigt for spillerne, men (hvad de matematiske formler ikke kan vide, men vi ved), så vil ingen spillere nogensinde spille et spil, hvor de skal dele deres gevinst med 10.000 andre.

Vi vil nu se på problemstillingen, hvor banken har lagt et loft $M=2^{N_0}$ over gevinsten svarende til, at spilleren slår krone N_0 gange i træk eller mere, så er gevinsten M.
Vi beregner derfor igen middelgevinsten med de nye spilleregler:

$$E(G)=\sum_{N=0}^{\infty}P(N)G(N)=\sum_{N=0}^{N_0-1}P(N)G(N)+\sum_{N=N_0}^{\infty}P(N)(M-S)$$

$$E(G)=\sum_{N=0}^{N_0-1}(\tfrac{1}{2})^{N+1}(2^N-S)+(M-S)\sum_{N=N_0}^{\infty}(\tfrac{1}{2})^{N+1}$$

$$=\tfrac{1}{2}N_0-\tfrac{1}{2}S\frac{1-(\frac{1}{2})^{N_0}}{1-\frac{1}{2}}+(M-S)(\tfrac{1}{2})^{N_0+1}\frac{1}{1-\frac{1}{2}}$$

$$E(G)=\tfrac{1}{2}N_0-S(1-(\tfrac{1}{2})^{N_0})+(M-S)(\tfrac{1}{2})^{N_0}$$

$$E(G)=\tfrac{1}{2}N_0-S+M(\tfrac{1}{2})^{N_0}$$

Indsætter man nu $M=2^{N_0}$ finder man: $E(G)=\tfrac{1}{2}N_0-S+1$.

Resultatet er (overraskende) simpelt. Hvis et Casino skal sikre sig en gevinst (i det lange løb), skal det altså blot sikre, at $E(G)<0$, altså at $S>\tfrac{1}{2}N_0+1$.

Hvis banken har et loft svarende til $N_0=16$, så $M=2^{16}=65.536$ jetoner, så skal indsatsen $S>8+1$ jetoner for at spillet er i Casino favør.

Det det kan forekomme at være en ret beskeden indsats, men sandsynligheden for at spilleren vinder maksimum M er imidlertid tilsvarende ringe, som allerede udregnet ovenfor.

$$P(M)=\sum_{N=N_0}^{\infty}(\tfrac{1}{2})^{N+1}=(\tfrac{1}{2})^{N_0+1}\frac{1}{1-\frac{1}{2}}=(\tfrac{1}{2})^{N_0}=\frac{1}{2^{16}}\quad for\quad N_0=16$$

Betingelsen for at en spiller får en gevinst, ved at slå N krone i træk er derfor:

$$S < 2^N \iff N > \log_2 S .$$

Sandsynligheden for dette er angivet ovenfor, hvor vi blot skal erstatte N_0 med N.

$$P(S < 2^N) = (\tfrac{1}{2})^N = \tfrac{1}{S} \quad \text{(I det sidste udtryk har vi sat } N = log_2 S\text{).}$$

Et bemærkelsesværdigt simpelt resultat, der imidlertid ikke er så attraktivt for en spiller, hvis S > 10 enheder. Hvis vi ønsker at finde sandsynligheden for at få en gevinst, der overstiger det dobbelte af indsatsen, har vi altså: $2S < 2^N \iff N - 1 > \log_2 S$, som giver sandsynligheden

$$P(S < 2^{N-1}) = (\tfrac{1}{2})^{N-1} = \tfrac{1}{2S} ,$$

og sådan fremdeles.